Claus O. Wilke
Digital Life Laboratory, Mail Code 136-93,
Caltech, Pasadena, CA 91125
wilke@caltech.edu
(Printed: June 29, 2001)

We introduce a model of molecular evolution in which the fitness of an individual depends both on its own and on the parent's genotype. The model can be solved by means of a nonlinear mapping onto the standard quasispecies model. The dependency on the parental genotypes cancels from the mean fitness, but not from the individual sequence concentrations. For finite populations, the position of the error threshold is very sensitive to the influence from parent genotypes. In addition to biological applications, our model is important for understanding the dynamics of self-replicating computer programs.

PACS numbers: 87.23.Kg

Simple models of asexual evolution, such as the quasispecies model, typically assume that the fitness of an organism is a function of only its genotype and the environment. This allows for the analysis of evolution in static [1–7] or variable [8–10] environments, many to one mappings from genotype to phenotype (neutrality) [11–13], and phenotypic plasticity [14]. These models disregard the potential influence of the mother (or parent, in general) on the organism's phenotype. This influence comes about because in addition to the genetic material, a wealth of proteins and other substances is transferred from mother to child. In the context of sexual reproduction, the influence of the mother on a child's phenotype is usually called a maternal effect. A classic example is that of the snail *Partula suturalis* [15], for which the directionality in the coiling of the shells is determined by the genotype of the mother of an organism, rather than the organism's own genotype. Maternal effects are not exclusive to sexually reproducing organisms, however, they can be observed in simple asexual organisms as well. In bacteria, for example, the fitness of a strain in a given environment may depend on the environment that was experienced by the strain's immediate ancestors [16,17].

Here, our objective is to define and study a model of the evolution of asexual organisms that takes such maternal effects into account. We assume that the fitness of an organism is given by the product of two quantities a and b, where a depends solely on the genotype of the mother of the organism, and b depends solely on the organism's own genotype. Since we need to keep track of the abundances of all possible mother/child combinations, we need n^2 concentration variables if we distinguish n different genotypes in our model. In the following, we denote by x_{ij} the concentration of organisms of genotype j descended from genotype i, and by q_{ji} the probability that a genotype i mutates into genotype j. The time evolution of the x_{ij} is then, in analogy to the quasispecies model,

$$\dot{x}_{ij}(t) = \sum_k a_k b_i q_{ji} x_{ki}(t) - f(t) x_{ij}(t)\,, \tag{1}$$

with $f(t) = \sum_{i,j} a_i b_j x_{ij}(t)$. The function $f(t)$ gives the average fitness of the population at time t. In principle, this model can be solved by diagonalizing a $n^2 \times n^2$ matrix. However, since even for relatively short genetic sequences the number of possible genotypes n is enormous, this direct method is very cumbersome. Fortunately, a simple transformation exists that reduces the above problem to one in which the diagonalization of a $n \times n$ matrix is sufficient. Namely, if we introduce variables x_i such that

$$x_{ij} = b_i q_{ji} x_i \Big/ \Big(\sum_k b_k x_k \Big)\,, \tag{2}$$

then, after inserting Eq. (2) into Eq. (1), we obtain in the steady state ($\dot{x}_{ij} = 0$)

$$\tilde{f} x_i = \sum_j a_j b_j q_{ij} x_j \tag{3}$$

with $\tilde{f} = \sum_j a_j b_j x_j$. The inverse transformation, which converts Eq. (3) back into the right-hand-side of Eq. (1), can be achieved with

$$x_i = \Big(\sum_j a_j x_{ji} \Big) \Big/ \Big(\sum_{j,k} a_j x_{jk} \Big)\,. \tag{4}$$

Therefore, Eq. (3) is fully equivalent to the steady state of Eq. (1). This leads to an interesting conclusion. Note that Eq. (3) is simply the steady state equation of the quasispecies model if we assume that genotypes j replicate with replication rate $c_j = a_j b_j$ and mutate into genotypes i with q_{ij}, while x_i gives the relative concentration of genotype i. Consequently, the model with maternal effects is mathematically equivalent to the standard quasispecies model. Moreover, with the aid of Eq. (2) and Eq. (4), it can be shown that $f(t \to \infty) = \tilde{f}$. Therefore,

the average fitness in both models is the same; the maternal effects drop out of the expression for the average fitness.

While the average fitness depends only on the values of c_i, the individual sequence concentrations actually depend on the details of the maternal effects. In particular, the total amount of sequences of a given genotype i is not identical to the corresponding value x_i in the standard quasispecies model ($\sum_j x_{ji} \neq x_i$ in general). From Eq. (2), we see that for every given mutation matrix q_{ij}, we can suppress any sequence concentration to as small a level as we please, by reducing the corresponding b_i and holding the other b_j constant ($\lim_{b_i \to 0} x_{ij} = 0$). Enhancement of sequence concentrations is also possible, although there exists an upper bound that cannot be exceeded. The upper bound is given by the matrix element q_{ji} ($\lim_{b_i \to \infty} x_{ij} = q_{ji}$). Its existence is easy to understand: by changing b_i, we can only jointly manipulate the concentrations of *all* sequences descended from genotype i. The ratio between different genotypes j descended from i is always fixed, and it is determined by the matrix elements q_{ji}. At most, the sum over all descendants from all genotypes i in the population can be one, $\sum_j x_{ij} = 1$, which implies $x_{ij} = q_{ji}$.

We will now classify the different types of maternal effects. If all $a_i = 1$, such that $b_i = c_i$, no maternal effects are present, and we obtain the normal sequence concentrations from the quasispecies model. We will refer to this situation as the *neutral* case. In order to classify all non-neutral situations, we compare concentrations of those sequences that are true copies of their parents (this is the only reasonable way, given that *all* genotypes descended from the same ancestor i scale identically with b_i). If the concentration x_{ii} of a sequence with large c_i is *reduced*, while the concentration x_{jj} of some other sequence with smaller c_j is *enhanced*, we will speak of *positive* maternal effects. Likewise, if the sequence concentration of a faster replicating sequence is *enhanced*, at the expense of some slower replicating sequence, we will speak of *negative* maternal effects. In short, positive maternal effects promote slower replicators, and negative maternal effects promote faster replicators. We refer to the above classification as the *direction* of the maternal effects. Likewise, the *strengh* of the maternal effects indicates the degree to which a system deviates from the neutral case (*weak* maternal effects show only a small deviation, *strong* maternal effects show a large deviation from the neutral case).

With Eq. (2) we can solve the model, as long as there exists an analytical solution for the corresponding quasispecies landscape. This means that solutions for multiplicative landscapes [18,19], the single peak landscape [20], and certain spin-glass landscapes [21–23] are readily available. In the following, we discuss the well-known example of the sharp single-peak landscape [24]. We assume that the genetic sequences are binary of length ℓ. Moreover, we assume a uniform copy fidelity q per digit. The sequence $000\ldots0$ may replicate (in the absence of mutations) with rate $c_0 = a_0 b_0$. We will refer to this sequence as the master sequence. Let all other sequences replicate with $c_1 = a_1 b_1 \ll c_0$. If ℓ is large, we may neglect back-mutations onto the master sequence, in which case it is sufficient to keep track of the total concentration of all sequences off the peak in a single variable, x_1. The mutation matrix q_{ij} is then a 2×2 matrix with the elements $q_{00} = q^\ell$, $q_{10} = 1 - q^\ell$, $q_{01} = q_{10}/(2^\ell - 1)$, and $q_{11} = 1 - q_{01}$ (the elements q_{01} and q_{01} are approximated). In the standard quasispecies model, the equilibrium concentration of the master sequence x_0 is given by

$$x_0 = (c_0 q_{00} - c_1)/(c_0 - c_1)\,, \tag{5}$$

and x_1 likewise as $x_1 = 1 - x_0$. The average fitness follows as

$$f = \begin{cases} c_0 q_{00} & \text{for } x_0 \geq 0, \\ c_1 & \text{else.} \end{cases} \tag{6}$$

Now, for the sequence concentrations with maternal effects, we obtain from Eq. (5) in conjunction with Eq. (2)

$$x_{0i} = b_0 q_{i0}(c_0 q_{00} - c_1)/\Lambda\,, \tag{7}$$

$$x_{1i} = b_1 q_{i1}(c_0 - c_0 q_{00})/\Lambda\,, \tag{8}$$

with $\Lambda = (b_0 - b_1)c_0 q_{00} + b_1 c_0 - b_0 c_1$ and $i = 0, 1$.

Figure 1 displays the four sequence concentrations x_{00}, x_{01}, x_{10}, x_{11} of the above defined landscape, for positive, negative, and neutral maternal effects. We see that indeed the maternal effects result in a significant shift in the sequence concentrations. The concentration x_{11} (shown in the lower right of Fig. 1), e.g., encompasses almost the complete population for positive maternal effects at an error rate of about 0.04, while it constitutes less than 20% in the case of the negative maternal effects for the same error rate.

The potential shift in the individual sequence concentrations has important implications for finite populations. When the concentration of a sequence (as predicted for an infinite population) approaches the inverse of the population size, that sequence will most certainly be lost through sampling fluctuations. In the case of the master sequence, this effect is responsible for the shift of the error catastrophe towards smaller error rates for finite populations in the ordinary quasispecies model [25–28]. Now, since the concentration of the master sequence can be made arbitrarily small with suitable maternal effects, it follows that the error threshold can be shifted. This effect is illustrated in Fig. 2 for a population size of $N = 1000$, for which the error transition in the normal quasispecies model (as represented by the 'neutral' case) is already significantly shifted. Positive maternal effects increase this shift by a fair amount, while negative maternal effects can almost completely counteract

the finite population effect, and move the error transition very close to the infinite population limit. Besides the shift in the error transition, Fig. 2 shows that the average fitness is indeed unaffected by strength and/or direction of the maternal effects, as all three curves lie exactly on the infinite population result for error rates below the respective error transitions.

We have seen above that the mean fitness in the population is not influenced by the existence of maternal effects. Since selection acts only on the average fitness [29,30], it follows that the maternal effects cannot be under selective pressure. In order to verify this, we have performed simulations in which strength and direction of the maternal effects were allowed to evolve. To each sequence in the population, we added an inheritable variable z. On reproduction, the offspring received a value $z' = z + dz$, where dz was a normally distributed random variable. For the master sequence, z was then mapped into a_0 and b_0 via $a_0 = (\alpha+z)/\alpha$ for $z > 0$, $a_0 = \alpha/(\alpha+z)$ for $z \leq 0$, and $b_0 = 1/a_0$, with α defining the scale between z and a_0, b_0. For a_1 and b_1, the value of z was ignored. Figure 3 shows a typical simulation run in such a system. We chose $N = 1000$ and $1-q = 0.06$, so that the population was below the error threshold in the absence of maternal effects, and we initialized all sequences in the population to $z = 0$. Over the course of a simulation, the z values drift randomly, which can be seen in increasing and diminishing fluctuations about the average fitness. When the average z drifts below zero, the fluctuations decrease, because $z < 0$ corresponds to negative maternal effects, which shift the population away from the error threshold. When the average z drifts above zero, on the other hand, the fluctuations increase. If there is no upper limit to z, the fluctuations will eventually grow so large that the population is pushed over the error threshold. In Fig. 3, this happend around generation 5400.

The model we have introduced in this paper oversimplifies the situation for bacteria, where substances can remain in the cytoplasm for several generations, such that not only the parent, but also the grand- and the grand-grand-parent etc. have an influence on the phenotype of an individual. However, it is an exact description of the dynamics of the digital organisms (self-replicating and evolving computer programs) of the Avida system, which has been used extensively in experimental evolution research [30–35]. The replication rate of these digital organisms is the ratio between the number of instructions per unit time that they can execute [the speed of their central processing unit (CPU)] and the number of instructions they have to execute in order to produce a new offspring (length of the gestation cycle). The CPU speed depends on the number and type of logical operations that these organisms perform in addition to their replicatory activity (the more logical operations an organism performs, the faster its CPU will operate). Since the logical operations an organism can perform are only known *a posteriori*, these organisms obtain their initial CPU speed from their parent organism. The CPU speed corresponds thus to the parameter a of the present work, and the length of the gestation cycle to the inverse of the parameter b. Therefore, we have shown in that a quasispecies description of the digital organisms is indeed justified, as was proposed in [30]. Also, our model might lead to a detailed quantitative description of the dynamics of digital organisms in future work.

This work was supported by the National Science Foundation under Contract No. DEB-9981397. The author thanks Chris Adami for carefully reading this manuscript.

[1] M. Eigen and P. Schuster, *The Hypercycle—A Principle of Natural Self-Organization* (Springer-Verlag, Berlin, 1979).
[2] M. Eigen, J. McCaskill, and P. Schuster, Adv. Chem. Phys. **75**, 149 (1989).
[3] E. van Nimwegen, J. P. Crutchfield, and M. Mitchell, Theoretical Computer Science **229**, 41 (1999).
[4] E. van Nimwegen and J. P. Crutchfield, Bull. Math. Biol. **62**, 799 (2000).
[5] A. Prügel-Bennett and J. L. Shapiro, Phys. Rev. Lett. **72**, 1305 (1994).
[6] M. Rattray and J. L. Shapiro, J. Phys. A: Math. Gen. **29**, 7451 (1996).
[7] A. Rogers and A. Prügel-Bennett, Theor. Pop. Biol. **57**, 121 (2000).
[8] M. Nilsson and N. Snoad, Phys. Rev. Lett. **84**, 191 (2000).
[9] C. O. Wilke and C. Ronnewinkel, Physica A **290**, 475 (2001).
[10] C. O. Wilke, C. Ronnewinkel, and T. Martinetz, Phys. Rep. **349**, 395 (2001).
[11] M. A. Huynen, P. F. Stadler, and W. Fontana, Proc. Natl. Acad. Sci. USA **93**, 397 (1996).
[12] C. Reidys, C. V. Forst, and P. Schuster, Bull. Math. Biol. **63**, 57 (2001).
[13] E. van Nimwegen, J. P. Crutchfield, and M. Huynen, Proc. Natl. Acad. Sci. USA **96**, 9716 (1999).
[14] L. W. Ancel and W. Fontana, J. Exp. Zoology **288**, 242 (2000).
[15] R. Dawkins, *The Extended Phenotype* (W. H. Freeman and Company, Oxford, 1982).
[16] A. M. Leroi, A. F. Bennett, and R. E. Lenski, Proc. Natl. Acad. Sci. USA **91**, 1917 (1994).
[17] R. E. Lenski *et al.*, Mol. Ecol. **3**, 127 (1994).
[18] D. S. Rumschitzki, J. Math. Biol. **24**, 667 (1987).
[19] G. Woodcock and P. G. Higgs, J. theor. Biol. **179**, 61 (1996).
[20] S. Galluccio, Phys. Rev. E **56**, 4526 (1997).
[21] I. Leuthäusser, J. Stat. Phys. **48**, 343 (1987).
[22] P. Tarazona, Phys. Rev. A **45**, 6038 (1992).

[23] S. Franz, L. Peliti, and M. Sellitto, J. Phys. A: Math. Gen. **26**, L1195 (1993).
[24] J. Swetina and P. Schuster, Biophys. Chem. **16**, 329 (1982).
[25] M. Nowak and P. Schuster, J. theor. Biol. **137**, 375 (1989).
[26] T. Wiehe, E. Baake, and P. Schuster, J. theor. Biol. **177**, 1 (1995).
[27] D. Alves and J. F. Fontanari, Phys. Rev. E **57**, 7008 (1998).
[28] P. R. A. Campos and J. F. Fontanari, J. Phys. A **32**, L1 (1999).
[29] P. Schuster and J. Swetina, Bull. Math. Biol. **50**, 635 (1988).
[30] C. O. Wilke *et al.*, Nature (2001), in press.
[31] C. Adami, *Introduction to Artificial Life* (Springer, New York, 1998).
[32] R. E. Lenski, C. Ofria, T. C. Collier, and C. Adami, Nature **400**, 661 (1999).
[33] C. Adami, C. Ofria, and T. C. Collier, Proc. Natl. Acad. Sci. USA **97**, 4463 (2000).
[34] D. Wagenaar and C. Adami, in *Proc. of Artificial Life VII*, edited by M. A. Bedau, J. McCaskill, N. Packard, and S. Rasmussen (MIT Press, ADDRESS, 2000), pp. 216–220.
[35] C. Ofria and C. Adami, in *Evolution as Computation*, edited by L. Landweber and E. Winfree (Springer, New York, 2001), p. 167.

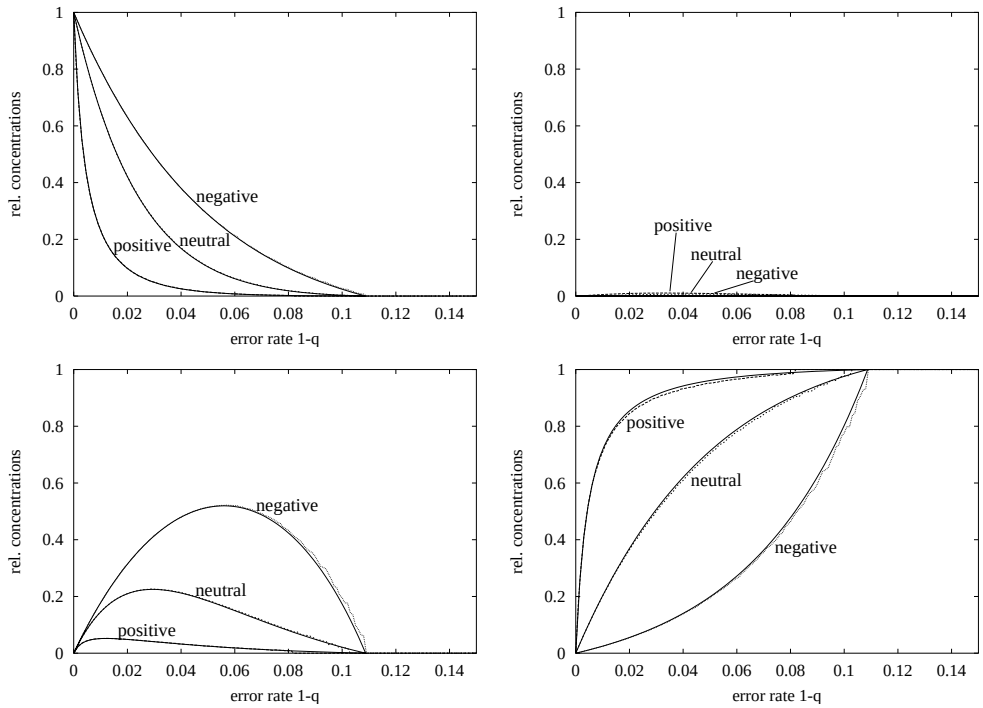

FIG. 1. Relative sequence concentrations vs. error rate $1-q$. From left to right, and from top to bottom, we display x_{00}, x_{10}, x_{01}, x_{11}. Solid lines are the analytical predictions Eqs. (7), (8), dashed lines stem from simulations with $N = 10000$ sequences of length $l = 20$. The parameters of the fitness landscapes were $c_0 = 10$ and $c_1 = 1$, with $b_0 = 0.1$, $b_1 = 1$ (positive); $b_0 = 1$, $b_1 = 1$ (neutral); $b_0 = 1$, $b_1 = 0.1$ (negative).

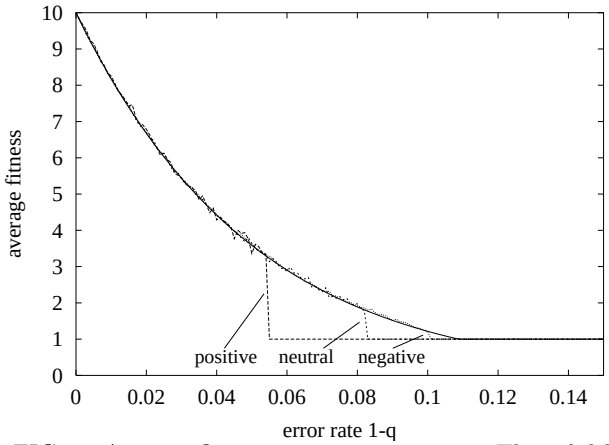

FIG. 2. Average fitness vs. error rate $1 - q$. The solid line represents Eq. (6), and the dashed lines stem from simulations with $N = 1000$ sequences of length $l = 20$. The fitness landscapes were identical to Fig. 1.

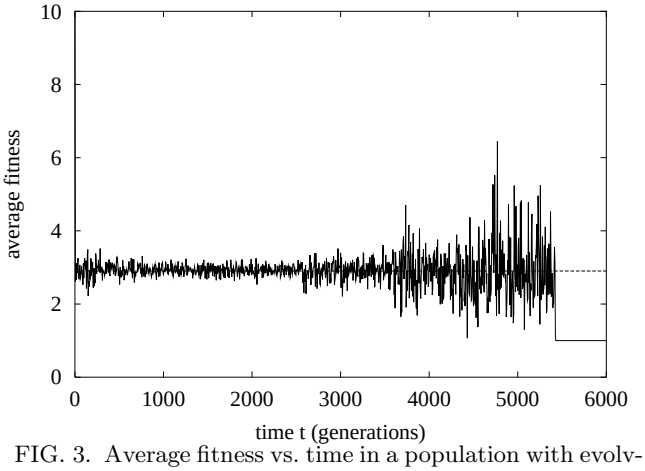

FIG. 3. Average fitness vs. time in a population with evolving maternal effects. The dashed line indicates the infinite population result [Eq. (6)]. The population consisted of $N = 1000$ sequencees of $l = 20$, the error rate was $1-q = 0.06$, and the landscape was defined by $c_0 = 10$, $c_1 = 1$. The scale parameter α was set to $\alpha = 10$.